# Continuously signal-adaptive filterbank for high-quality perceptual audio coding

Herre, J. Johnston, J.D.

*This paper appears in:* **Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on**

Meeting Date: 10/19/1997 - 10/22/1997
Publication Date: 19-22 Oct. 1997
Location: New Paltz, NY USA
On page(s): 4 pp.
Reference Cited: 15
Number of Pages: 298
Inspec Accession Number: 5862506

**Abstract:**
Historically, the choice of the optimum filterbank has been the subject of much research and discussion in the development of perceptual audio coders. Desirable properties of a good filterbank include both a good extraction of the signal's redundancy and effective utilization of that redundancy while maintaining control over perceptual demands. Often, there is a conflict between the use of perceptual constraints and the redundancy extraction, in that a filterbank with good resolution in both time and frequency is needed. Recently, a method for performing temporal noise shaping (TNS) of the error signal of a perceptual audio coder has been proposed, providing control over both the time and frequency structure of the coding noise. This paper focuses on the core part of the scheme, forming a continuously adaptive filterbank, and discusses its theoretical background, properties and limitations

**Index Terms:**
adaptive filters adaptive signal processing audio coding band-pass filters coding errors filtering theory noise signal resolution coding noise continuously signal-adaptive filterbank error signal frequency resolution high-quality perceptual audio coding optimum filterbank perceptual audio coder perceptual audio coders perceptual constraints perceptual demands control signal

# CONTINUOUSLY SIGNAL-ADAPTIVE FILTERBANK FOR HIGH-QUALITY PERCEPTUAL AUDIO CODING

*Jürgen Herre[1) and James D. Johnston[2)*

[1) Fraunhofer Institute Integrated Circuits FhG-IIS, hrr@iis.fhg.de

[2) AT&T Laboratories - Research, jj@research.att.com

## ABSTRACT

Historically, the choice of the optimum filterbank has been the subject of much research and discussion in the development of perceptual audio coders. Desirable properties of a good filterbank include both a good extraction of the signal's redundancy and effective utilization of that redundancy while maintaining control over perceptual demands. Often, there is a conflict between the use of perceptual constraints and the redundancy extraction, in that a filterbank with good resolution in both time and frequency is needed. Recently, a method for performing temporal noise shaping (TNS) of the error signal of a perceptual audio coder has been proposed, providing control over both the time and frequency structure of the coding noise. This paper focuses on the core part of the scheme, forming a continuously adaptive filterbank, and discusses its theoretical background, properties and limitations.

## 1. OPTIMUM FILTERBANKS FOR PERCEPTUAL AUDIO CODING

The time/frequency mapping scheme (analysis filterbank) is the central part of a perceptual audio coder (see figure 1) and is crucial for the coder's performance. Historically, the choice of the optimum filterbank has been the subject of much research and discussion in the development of perceptual audio coders [1] [2] [3] [4] [5] [6] [7] [8]. A summary of requirements for efficient filterbanks in a perceptual audio coder is given in [9]. Desirable properties of a good filterbank include both good extraction of the signal redundancy and utilization of signal irrelevancy. This can be described for three extrema as follows:

- For stationary or pseudo-stationary signals with many frequency components (harmonics) e.g. Harpsichord, it is essential to have a sufficiently large transform size to resolve the lines in the signal spectrum in order to extract the redundancy. Because of the stationary properties of the signal envelope at most, if not all frequencies, protection from time domain artifacts does not play a major role for these signals. Thus, a high-resolution uniform filterbank is necessary in such cases, and a penalty of 10-30 dB in loss of coding gain will be paid with a low-resolution filterbank.
- For transient signal types, e.g. castanets, the emphasis of the coding process is on the removal of irrelevance by optimally exploiting the masking properties of the human auditory system. Since a fine structure in frequency in the high frequency range is not available during transient signal portions, the optimum choice for such cases is a critical band filter structure. The inefficiency involved in using a high-resolution filterbank for such signals can amount to 70-80 dB.
- For signals like pulse trains, or those of a pitch-periodic nature, e.g. speech, the high frequency content is clustered in time around each pitch event. In such cases, the coding noise at high frequencies must follow the fine time structure of the signal, if the pitch period is longer than the main lobe of the transient response of the cochlear filter at the respective frequency. Using a uniform high-resolution filterbank for such signals can involve inefficiencies from a few dB up to about 15 dB for some critical signals. An alternative approach is to use a critical band filter structure as described in [1] and [5]. These filterbanks, however, have not considered the optimum combination of irrelevancy and redundancy extraction, and have focused on the perceptual demands at the expense of coding efficiency.

Obviously, the optimum filterbank for any input signal ranges somewhere in the continuum between these extreme cases depending on the spectral / temporal characteristics of the signal. Thus, the optimum filterbank needs to adapt to the characteristics of the signal, and to vary time and frequency resolution in a frequency and signal dependent fashion.

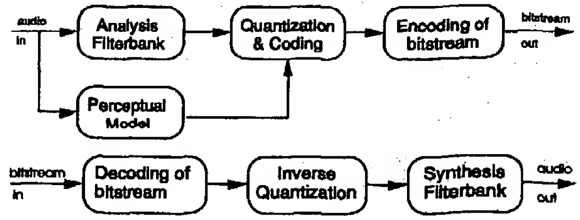


**Figure 1:** Generic block diagram of a perceptual audio coder

## 2. COMMONLY USED FILTERBANKS

In practice, because of the ease of implementation and computational complexity constraints, most coders use a uniformly spaced filterbank instead of the critical band structure also for coding of transient signal parts.

Examples of common coders with a uniform filterbank and a low number of filterbank channels (low frequency resolution) are ISO/MPEG-Audio Layers I and II [4] using a 32 band polyphase filterbank [7]. Some well-established coders with a high number of filterbank channels (high frequency resolution) are ISO/MPEG-Audio Layer III [4], PAC [10], AC-3 [11] and ISO/MPEG2 Advanced Audio Coding (AAC) [12] [13] using MDCT-based filterbanks [6]. In the latter coder family, a switched adaptation to the signal characteristics of the input signal is performed by a window switching operation [8] which allows to select a second (lower) frequency resolution for the coding of transient signal parts (see figure 2). Recently, also a coder has been presented with a switched time/frequency resolution achieved by switching between MDCT and wavelet based filterbanks [5].

In the described perceptual coders using window switching, however, no soft transition between extreme filterbank characteristics is possible, rather all signals are handled by "hard" switching between high and low frequency resolution or between a uniform and a non-uniform filterbank.

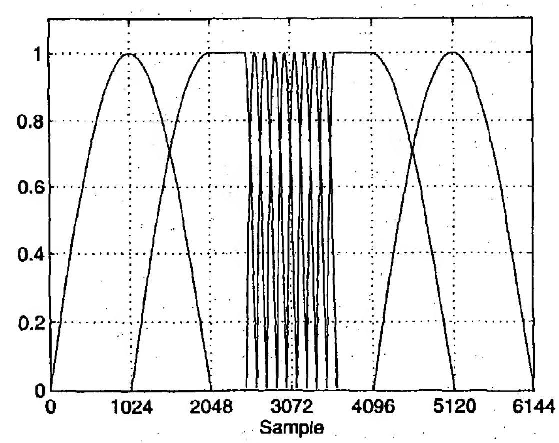


**Figure 2:** Principle of MDCT window switching

## 3. THE TNS FILTERBANK

The Temporal Noise Shaping (TNS) approach, as introduced in [14], is an extension of the standard scheme of a perceptual coder permitting the coder to exercise control over the temporal fine structure of the coding noise even within a filterbank window. The approach is based on the following considerations:

__Time / Frequency Duality Considerations:__
The concept of TNS is based upon the dual of the standard LPC analysis paradigm. It is well-known that signals with an "un-flat" spectrum can be coded efficiently either by directly coding spectral values ("transform coding") or by applying predictive coding methods to the time signal [Jay84*]. Consequently, the corresponding dual statement relates to the coding of signals with an "un-flat" time structure, i.e. transient signals. Efficient coding of transient signals can thus be achieved by either directly coding time domain values or by employing *predictive coding methods to the spectral data* by carrying out a prediction across frequency. In fact, it can be shown that due to the duality between time and frequency the amount of "prediction gain" (i.e. reduction of residual energy) reached is determined by the "unflatness" of the signal's temporal envelope in the same fashion that the Spectral Flatness Measure (SFM) is a measure of the reduction of residual energy available by LPC prediction.

__Noise Shaping by Predictive Coding:__
If an open-loop predictive coding technique is applied to *a time signal*, the quantization error in the final decoded signal is known to be *adapted in its Power Spectral Density* (PSD) to the PSD of the input signal (D*PCM, [15]). Dual to this, if predictive coding is *applied to spectral data over frequency*, the *temporal shape of the quantization error* signal will appear adapted to the temporal shape of the input signal at the output of the decoder. This effectively puts the quantization noise under the actual signal and in this way avoids problems of temporal masking, either in transient or pitched signals. This type of predictive coding of spectral data is therefore referred to as the "Temporal Noise Shaping" (TNS) approach.

A more rigorous derivation of these properties was published in [14] showing that the squared Hilbert envelope of a signal and the power spectral density constitute dual aspects in time and frequency domain.

__Implementation into a Perceptual Coder:__
The predictive encoding/decoding process over frequency can be realized easily by adding one block to the standard structure of a generic perceptual encoder and decoder. This is shown in figure 3. An additional block, "TNS Filtering" is inserted after the analysis filterbank performing an in-place filtering operation on the spectral values, i.e. replacing the target spectral coefficients (set of spectral coefficients to which TNS should be applied) with the prediction residual. This is symbolized by a "rotating switch circuitry in the figure. Both sliding in the order of increasing and decreasing frequency are possible.

Similarly, the TNS decoding process is done by inserting an additional block, "Inverse TNS Filtering", immediately before the synthesis filterbank (see figure 4). An inverse in-place filtering operation is performed on the residual spectral values so that the target spectral coefficients are replaced with the decoded spectral coefficients by means of the inverse prediction (all-pole) filter.

The TNS operation is signaled to the decoder via a dedicated part of the side information that includes a TNS on/off flag and the prediction filter data.

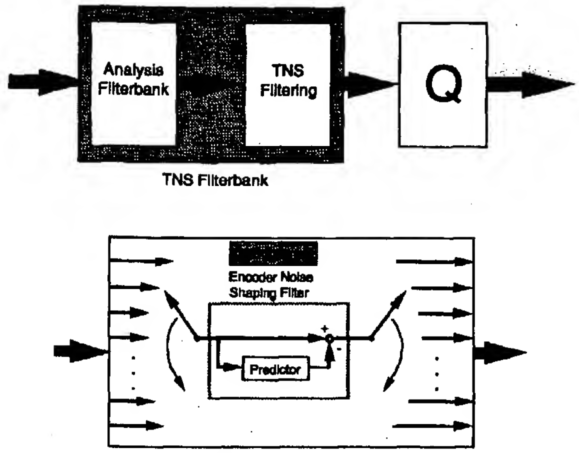


**Figure 3:** TNS processing in the encoder

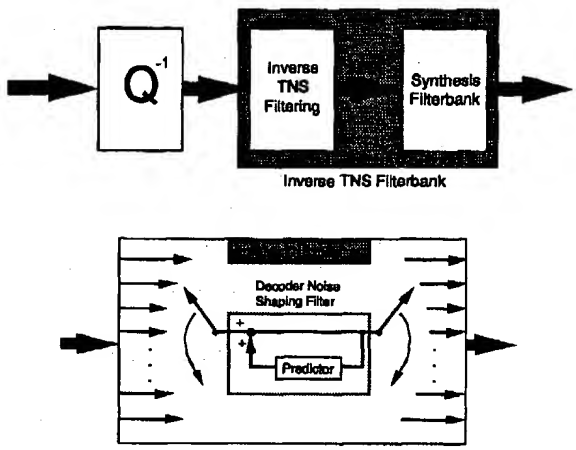


**Figure 4:** TNS processing in the decoder

While the derivation of the TNS approach was based on considerations of predictive coding methods, it is most instructive to interpret the combination of filterbank and prediction filter as a composite filterbank with a number of interesting properties.

# 4. PROPERTIES OF THE TNS FILTERBANK

## 4.1 Continuous Adaptation To Signal

In contrast to the hard switched filterbank schemes described previously, the Temporal Noise Shaping filterbank allows a continuous adaptation to the properties of the input signal in the following way:

- For signals with a considerable correlation between adjacent spectral coefficients (i.e. for signals with a very "unflat" envelope) the prediction filter will combine (convolve) these coefficients to calculate the prediction residual.
- In this way, frequency resolution will decrease and is traded adaptively in favor of temporal resolution. Note that the filterbank's increased temporal resolution is not represented by a number of timely subsequent spectral coefficients but by a multitude of coefficients of the same time instant corresponding to largely overlapping (widened) frequency bins.

Thus, the frequency (and time) resolution is adjusted adaptively to the input signal. This enables the interpretation of the combination of filterbank and adaptive prediction filter as a *continuously adaptive filterbank* as opposed to the classic "switched filterbank" approach. In fact, this type of adaptive filterbank dynamically provides a continuum in its behavior between a high-resolution filterbank (for stationary signals) and a low-resolution filterbank (for transient signals) and therefore approaches the requirements mentioned above for the optimum filterbank for a given input signal.

## 4.2 Time Domain Aliasing

In the previous chapters the discussion of the Temporal Noise Shaping filterbank was based on the notion of Fourier transform (and Discrete Fourier Transform DFT in the case of discrete spectral coefficients). In practice, the MDCT is preferred over the FFT and/or DCT in a modern transform coder for the reasons that it is both critically sampled and delivers an excellent coding efficiency.

It can be shown that the TNS filterbank provides a straightforward temporal noise shaping effect also for the known classic orthogonal block transforms like Discrete Fourier Transform (DFT) or Discrete Cosine or Sine Transform (DCT, DST). If the perceptual coder uses a critically subsampled filterbank with overlapping windows (e.g. an MDCT or any other filterbank based on Time Domain Aliasing Cancellation TDAC [6]) the resulting temporal noise shaping is also subject to the time domain aliasing effects inherent in this filterbank. For example, in the case of a MDCT one mirroring (aliasing) operation per window half takes place and the quantization noise appears mirrored (aliased) within the left and the right half of the window after decoding, respectively. Since the final filterbank output is obtained by applying a synthesis window to the output of each inverse transform and performing an overlap-add of these data segments, the undesired aliased components are attenuated depending on the analysis-synthesis window. Thus it is advantageous to choose a filterbank window that exhibits only a small overlap between subsequent blocks such that the temporal aliasing effect is minimized.

# 5. USING THE TNS FILTERBANK IN A PERCEPTUAL AUDIO CODER

By incorporating the TNS filterbank, a perceptual audio coder will benefit as follows:

- It permits for a better encoding of "pitch-based" signals such as speech which consist of a pseudo-stationary series of impulse-like signals without penalty in coding efficiency.

- The method reduces the peak bit demand of the coder for transient signal segments by exploiting irrelevancyby reducing the required pre-echo protections for such signals. As a side effect, the coder can stay longer in the preferred "long block" mode so that use of the less efficient "short block" mode can be reduced.

- The technique can be combined with other methods for addressing the temporal noise shaping problem such as block switching. Using temporal noise shaping it may, however, be possible to omit the need for a second coder mode (short block mode) leading to a simplified encoder / decoder structure.

- Since TNS processing can be applied either for the entire spectrum or for only part of the spectrum, the time-domain noise control can be applied in any necessary frequency-dependent fashion. In particular, it is possible to use several filters operating on distinct frequency (coefficient) regions, or to provide no TNS processing at some frequencies.

## 6. STANDARDIZATION

TNS has been adopted in the ISO/MPEG-2 Advanced Audio Coding (AAC) standard [12] [13] through the core experiment process. The results of the core experiment for the inclusion of the TNS filterbank showed an improvement for the critical "German Male Speech" test signal by about 0.9 points on the 5-grade ITU-R impairment scale. This improvement was due to mitigation of audible noise that was occurring between the pitch pulses in the speech by the TNS process. Depending on the profile, the TNS process as specified in the standard has a limit of 12 (low complexity profile) or 20 (main profile) poles for a block of 1024 frequency lines in the AAC standard and is activated according to signal demands.

## 7. CONCLUSIONS

A novel concept for a continuously signal-adaptive filterbank has been presented. This Temporal Noise Shaping filterbank technique works efficiently in the critical case of transient and "pitch-based" signals (e.g. speech) providing a noise shaping effect even within one block. The performance gain by using the TNS filterbank has been verified in the recent development process of the MPEG2-Audio AAC coder. Due to the general nature of the discussed issues, the generic principle of predictive coding "across frequency" may also have applications in different fields of perceptual coding (e.g. in image coding, addressing "edge effects").

## ACKNOWLEDGMENTS

A major part of the work on Temporal Noise Shaping has been done during Dr. Herre's term at Bell Laboratories. The authors wish to express their thanks to David Malah of the Technion Israel Institute for Technology and to Peter Kroon and Emre Telatar of Bell Laboratories for their valuable discussions about predictive coding methods and the mathematics of Hilbert envelopes, respectively.

## REFERENCES

[1] G. Theile, G. Stoll, M. Link: "Low Bit-Rate Coding of High-Quality Audio Signals. An Introduction to the MASCAM System", EBU Review-Technical No 230, August 1988, pp. 158-181.

[2] K. Brandenburg, J. D. Johnston: "Second Generation Perceptual Audio Coding: The Hybrid Coder", 88th. AES Convention, Montreux 1990, Preprint 2937

[3] K. Brandenburg, J. Herre, J.D. Johnston, Y. Mahieux, E.F. Schroeder: "ASPEC: Adaptive Spectral Perceptual Entropy Coding of High Quality Music Signals", 90th. AES Convention, Paris 1991, Preprint 3011

[4] K. Brandenburg, G. Stoll, Y.F. Dehéry, J.D. Johnston, L.v.d. Kerkhof, E.F. Schroeder: "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio", 92nd AES Convention, Wien 1992, Preprint 3336

[5] D. Sinha, J. D. Johnston: "Audio Compression at Low Bit Rates Using a Signal Adaptive Switched Filterbank", Proc. IEEE ICASSP, 1996, pp. 1053ff

[6] J. Princen, A. Johnson, A. Bradley: "Subband / Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation", IEEE ICASSP 1987, pp. 2161 - 2164

[7] J.H. Rothweiler: "Polyphase Quadrature Filters - a new Subband Coding Technique", IEEE ICASSP 1983, Boston, pp. 1280 - 1283

[8] B. Edler: "Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen", Frequenz, Vol. 43, pp. 252-256, 1989 (in German)

[9] J. D. Johnston: "Audio Coding with Filter Banks", pages 287-307 in: "Subband and Wavelet Transforms" by A. N. Akansu and M. J. T. Smith (editors), Kluwer Academic Publishers, Norwell 1996

[10] J. D. Johnston, A. J. Ferreira: "Sum-Difference Stereo Transform Coding", IEEE ICASSP 1992, pp. 569-571

[11] Mark Davis: "The AC-3 Multichannel Coder", 95th AES Convention, New York 1993, Preprint 3774

[12] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Oikawa: "MPEG-2 Advanced Audio Coding", 101st AES Convention, Los Angeles 1996

[13] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO 13818-7 "Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding"

[14] J. Herre, J. D. Johnston, "Enhancing the Performance of Perceptual Audio Coders by Using Temporal Noise Shaping (TNS)", 101st AES convention, Los Angeles 1996, Preprint 4384

[15] N. Jayant, P. Noll: "Digital Coding of Waveforms", Englewood Cliffs, NJ, Prentice-Hall, 1984